Data Management and Research Misconduct

Claudia Neuhauser, Ph.D., Interim VP for Research

Santi Thompson, M.A., M.L.I.S., Associate Dean, Research and Student Engagement, Libraries

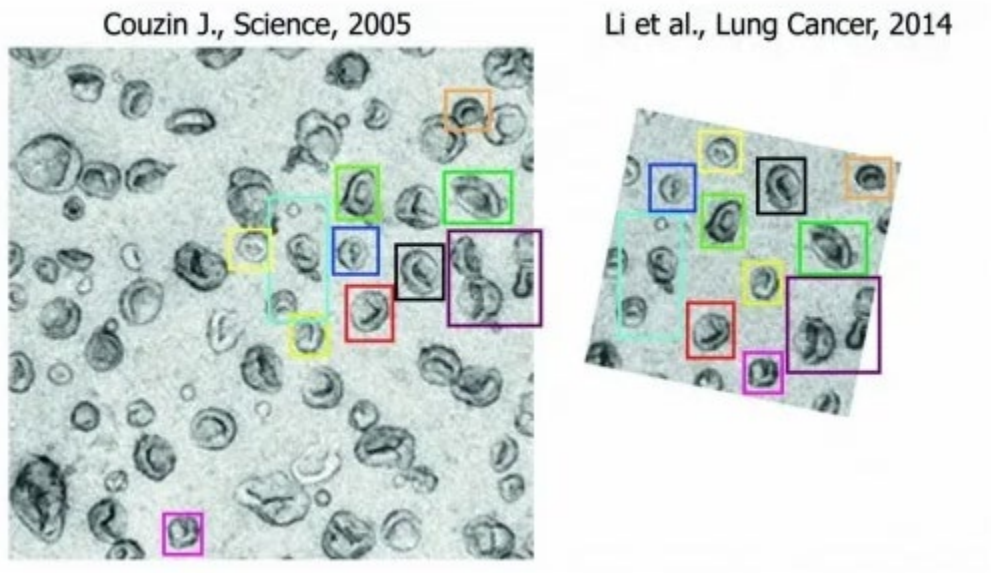
Topics

- What is Research Misconduct
- Image Manipulation
- What Drives Researchers to Engage in Research Misconduct
- Research Misconduct: Process
- A Case Study
- Data Management Best Practices

What is Research Misconduct?

Definition of Research Misconduct

- Federal government definition of research misconduct
  - *Fabrication:* Making up data and recording or reporting them
    - Duplicating survey data
  - *Falsification:* Manipulating research materials, equipment, or processes, or changing or omitting data or results
    - Image manipulation
  - *Plagiarism:* Appropriation of another person's ideas, processes, results, or words without giving appropriate credit



From: Cancer paper that doctored image from Science story earns retraction https://retractionwatch.com/2016/05/04/cancer-paper-that-doctored-image-from-science-story-earns-retraction/

Prevalence of Research Misconduct

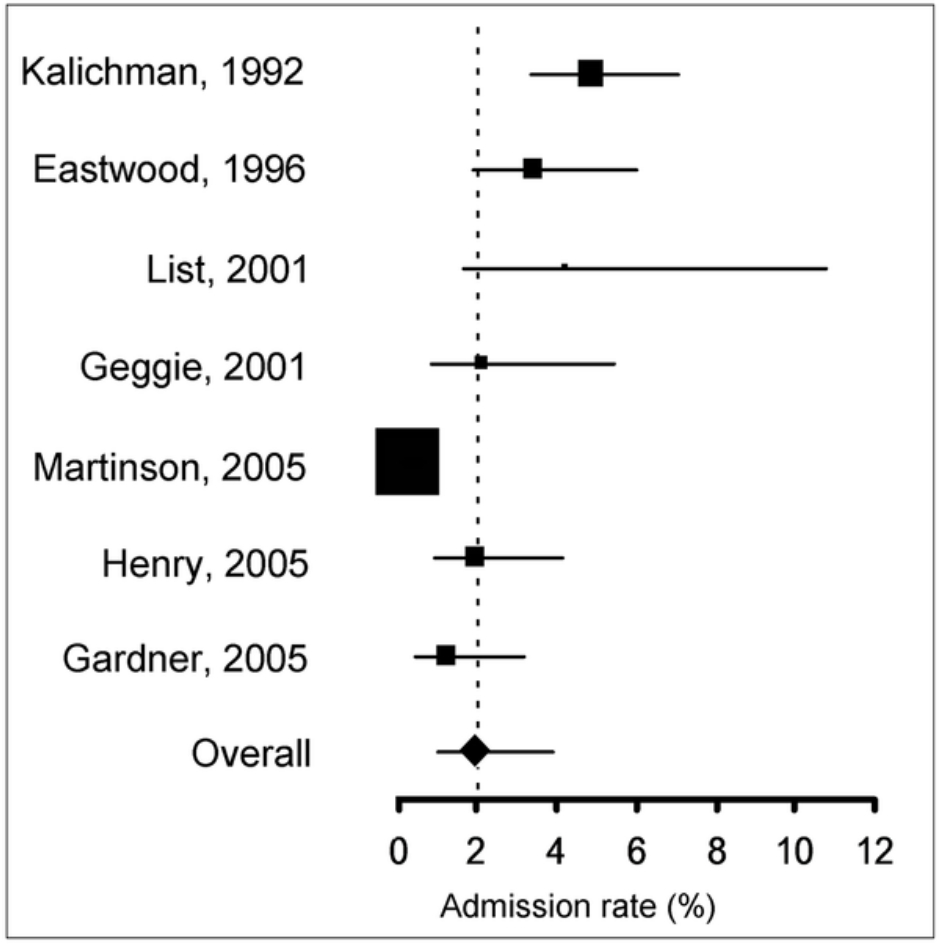


Figure 2. Forrest plot of admission rates of data fabrication, falsification and alteration in self reports.

"When explicitly asked if they ever fabricated or falsified research data or if they altered or modified results to improve the outcome (see Table S2, questions 1, 4, 6, 8, 10, 17, 26), between 0.3% and 4.9% of scientists replied affirmatively (N = 7, crude unweighted mean: 2.59%, 95%CI = 1.06–4.13)."

Fanelli D (2009) How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. PLOS ONE 4(5): e5738. https://doi.org/10.1371/journal.pone.0005738
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0005738

Examples of Research Misconduct

- Source: List of scientific misconduct incidents. Wikipedia. https://en.wikipedia.org/wiki/List_of_scientific_misconduct_incidents
  - Citation numbers below refer to this Wikipedia page. (Emphasis is ours.)
- Physics and Engineering
  - Jan Hendrik Schön (Germany, US), a researcher in the physics of semiconductors formerly employed by Bell Labs, forged results by **using the same data sets for different and unrelated experiments**.[235][236] Schön has had 32 of his publications retracted.[173]
- Chemistry
  - The independent misconduct of two chemists at the William A. Hinton State Laboratory Institute in Massachusetts caused the drug lab to be shut down and tens of thousands of criminal convictions for drug possession to be overturned. Annie Dookhan admitted to **faking test results** and adulterating samples to make them consistent with her desired results. Sonja Farak admitted to stealing samples and using them to get high herself.[201] The affairs were documented in the 2020 film *How to Fix a Drug Scandal*.
- Biology and biomedical sciences
  - Elias Alsabti (Iraq, US), was a medical practitioner who posed as a biomedical researcher. He **plagiarized** as many as 60 papers in the field of cancer research, many with non-existent co-authors.[11][12][13]
- Philosophy
  - Martin William Francis Stone, an Irish philosopher formerly at the Katholieke Universiteit Leuven, **plagiarized** in more than 40 publications.[223]

Questionable Research Practices (QRPs)

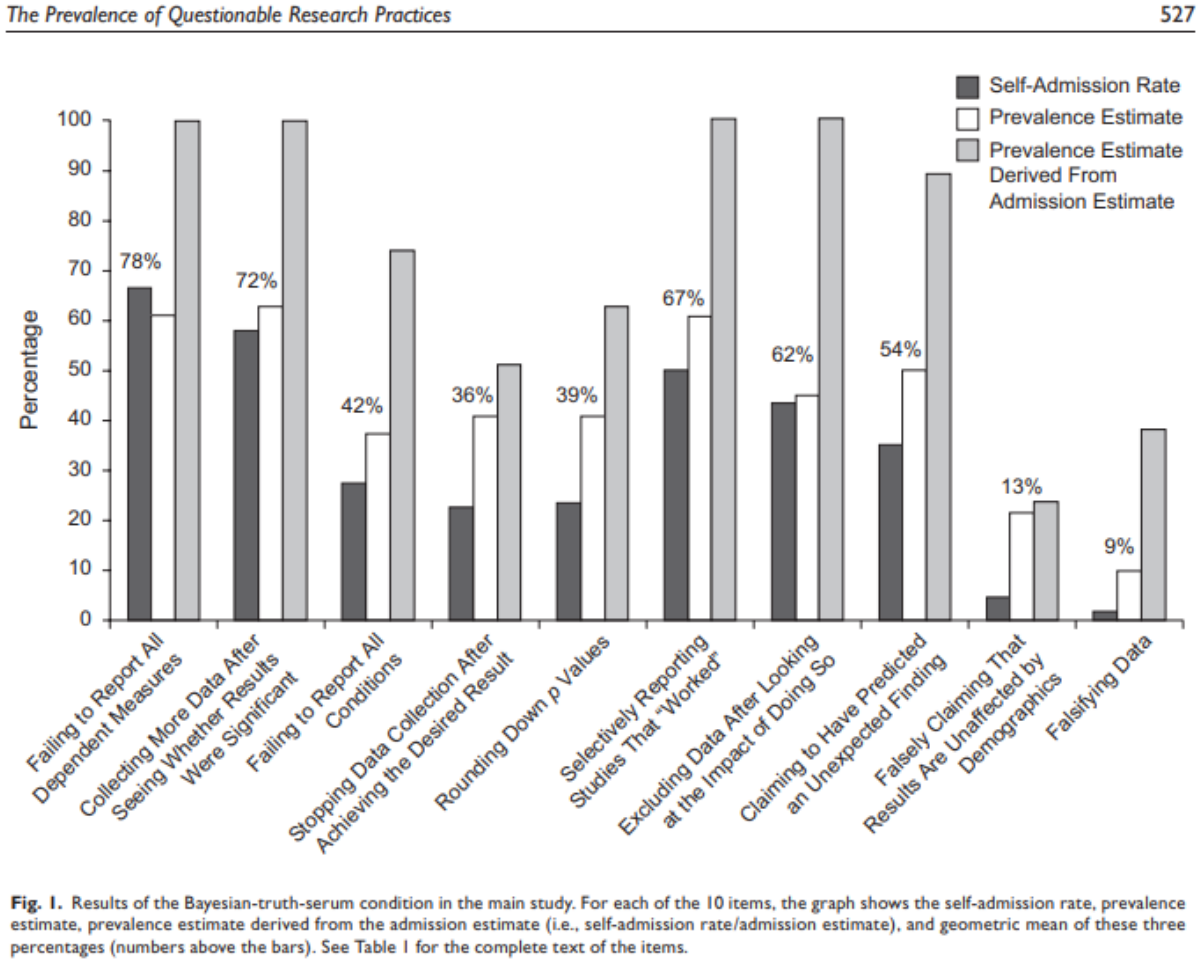


Fig. 1. Results of the Bayesian-truth-serum condition in the main study. For each of the 10 items, the graph shows the self-admission rate, prevalence estimate, prevalence estimate derived from the admission estimate (i.e., self-admission rate/admission estimate), and geometric mean of these three percentages (numbers above the bars). See Table 1 for the complete text of the items.

John, Leslie K., George Loewenstein, and Drazen Prelec. "Measuring the prevalence of questionable research practices with incentives for truth telling." Psychological science 23.5 (2012): 524-532.

Questionable Research Practices (QRPs)

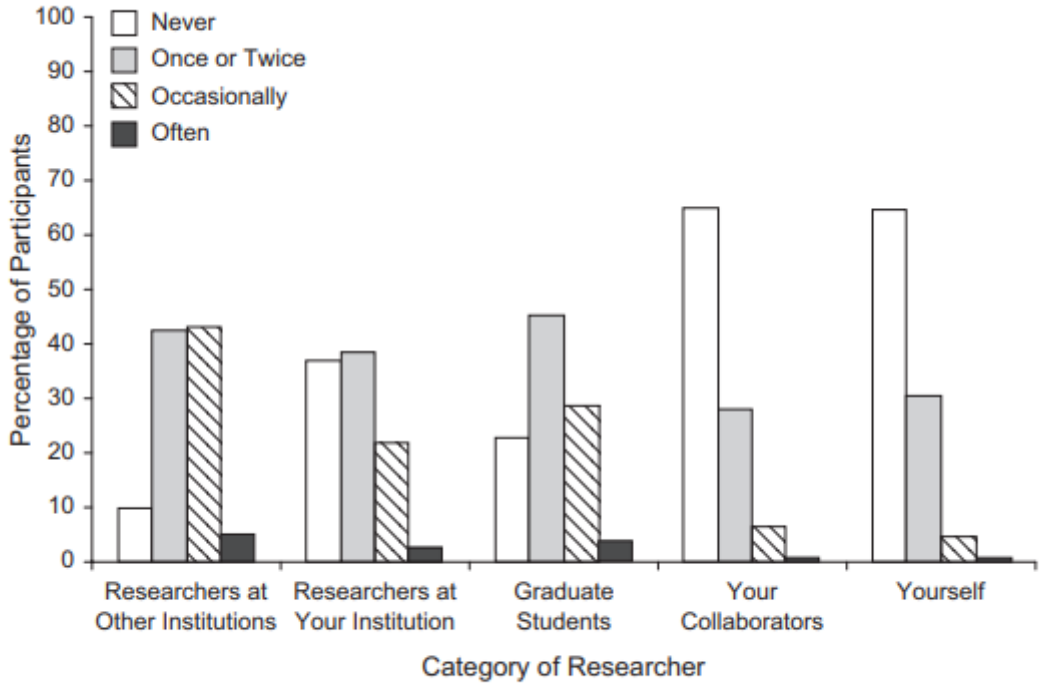


Fig. 2. Results of the main study: distribution of responses to a question asking about doubts concerning the integrity of the research conducted by various categories of researchers.

John, Leslie K., George Loewenstein, and Drazen Prelec. "Measuring the prevalence of questionable research practices with incentives for truth telling." Psychological science 23.5 (2012): 524-532.

Consequences

SCIENCEINSIDER | SCIENTIFIC COMMUNITY

Duke University settles research misconduct lawsuit for $112.5 million

Former university biologist who blew the whistle on data fabrication could get as much as one-third of total

25 MAR 2019 · BY SCIENCE NEWS STAFF

“Thomas alleged that Duke biologist Erin Potts-Kant—a co-author on numerous papers that are now retracted—included fraudulent data in 60 grant reports and funding applications to U.S. agencies. ‘Duke discovered the possible research misconduct in 2013 after [Potts-Kant] was fired for embezzling money from the university, which also occurred over the same period,’ the university noted in a statement released today. Potts-Kants ‘eventually pled guilty to two counts of forgery and paid restitution to Duke.’"

Consequences

Published: 01 July 2015

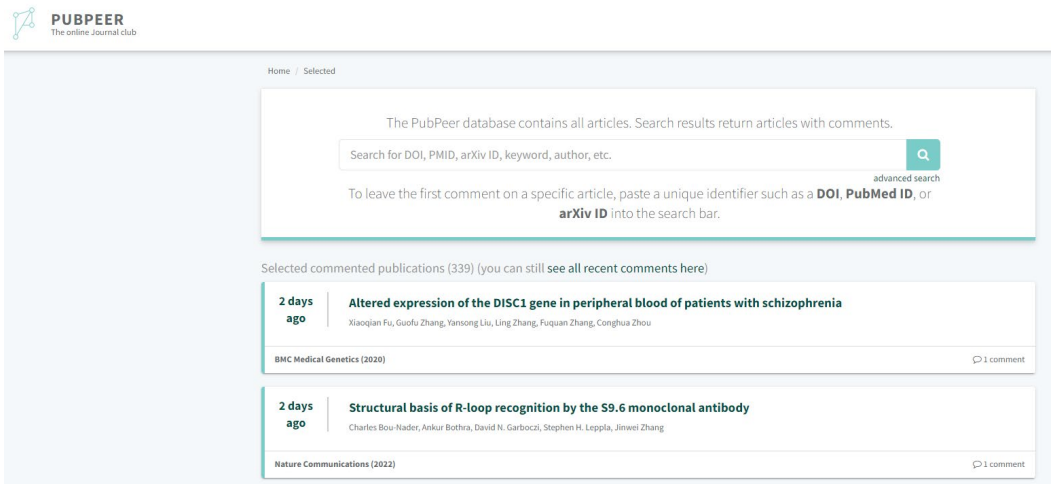
US vaccine researcher sentenced to prison for fraud

Sara Reardon

Nature **523**, 138–139 (2015) | Cite this article

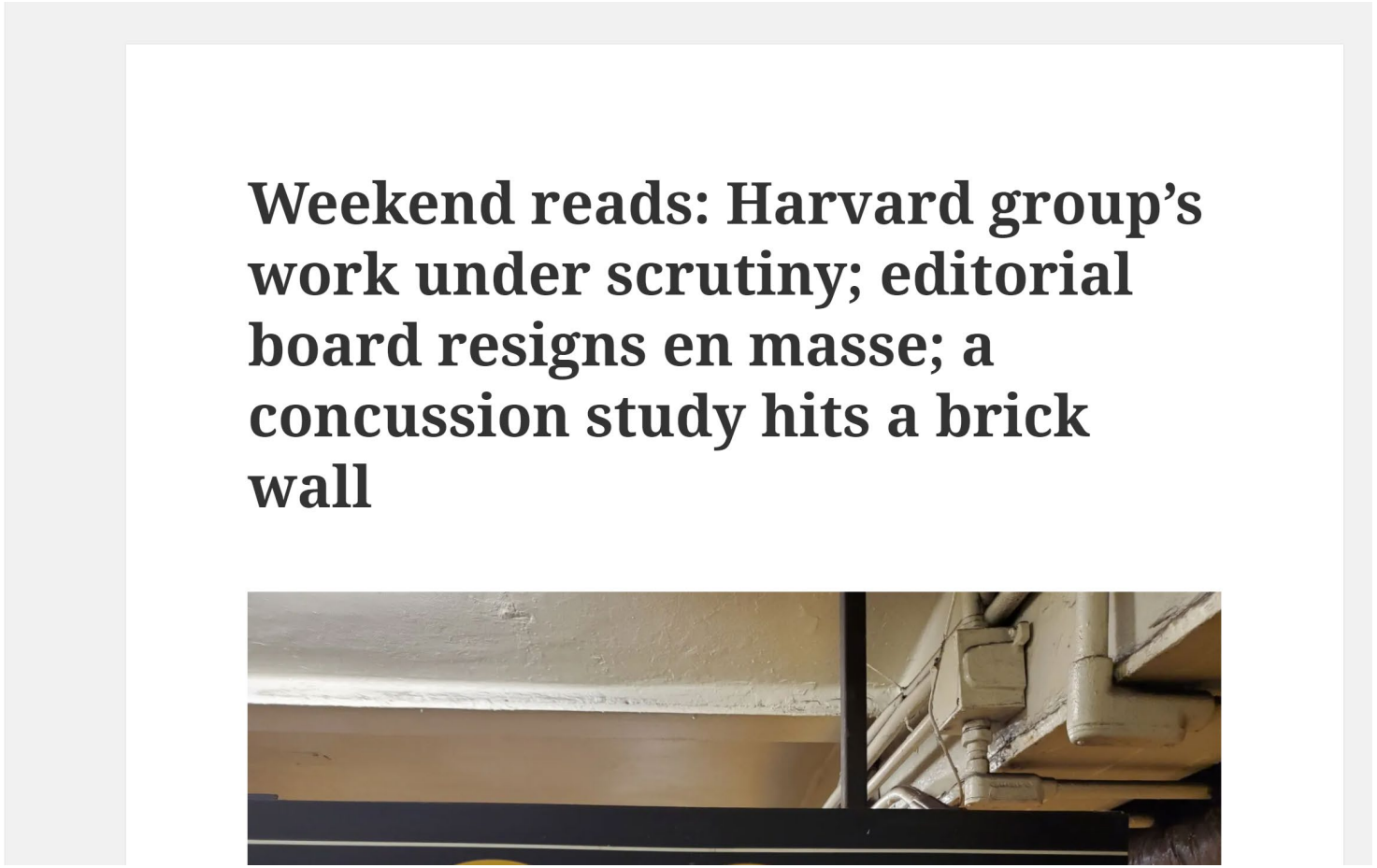
"Rare is the scientist who goes to prison on research misconduct charges. But on 1 July, Dong-Pyou Han, a former biomedical scientist at Iowa State University in Ames, was sentenced to 57 months for fabricating and falsifying data in HIV vaccine trials. Han has also been fined US $7.2 million and will be subject to three years of supervised release after he leaves prison."

PubPeer (https://pubpeer.com/)



Shen, Helen. "Meet this super-spotter of duplicated images in science papers." *Nature* 581.7807 (2020): 132-137. (https://www.nature.com/articles/d41586-020-01363-z; accessed on April 21, 2023)

Retraction Watch



https://retractionwatch.com/

Image Manipulation

Image Manipulation

- Why?
  - Inexperience, impatience, lack of resources to repeat the experiment, increase impact of publication
- What's acceptable?
  - Resizing and magnification that does not distort or alter the image
  - Adjusting brightness/contrast if applied to entire image
  - Annotation of image
- What's not acceptable?
  - Changing gray scale to color
  - Sharpening or over-saturating images
  - Reusing images for a different experiment or combining images from different experiments to present a single experiment
  - Eliminating background noise and not declaring it
  - Manipulation of signals/bands in blots
- Source: Joshi, Adita. Image manipulation: What's okay and not okay and who addresses it? editage insights. June 2022. (https://www.editage.com/insights/image-manipulation-whats-okay-and-not-okay-and-who-addresses-it; accessed on April 21, 2023)

How to Avoid Inappropriate Image Manipulation?

- Before capturing an image
  - Understand proper use of imaging instrument/software
  - Prepare clean samples (minimizes noise and maximizes resolution and sharpness)
  - Record all settings used to produce images for replicability
- After capturing an image
  - Preserve original; make changes only on copies
  - Save images as TIFF files (not JPEG)
  - Record all settings used to finalize the image for reproducibility
  - Mention any manipulations in the figure legend
- Source: Joshi, Adita. Image manipulation: What's okay and not okay and who addresses it? editage insights. June 2022. (https://www.editage.com/insights/image-manipulation-whats-okay-and-not-okay-and-who-addresses-it; accessed on April 21, 2023)

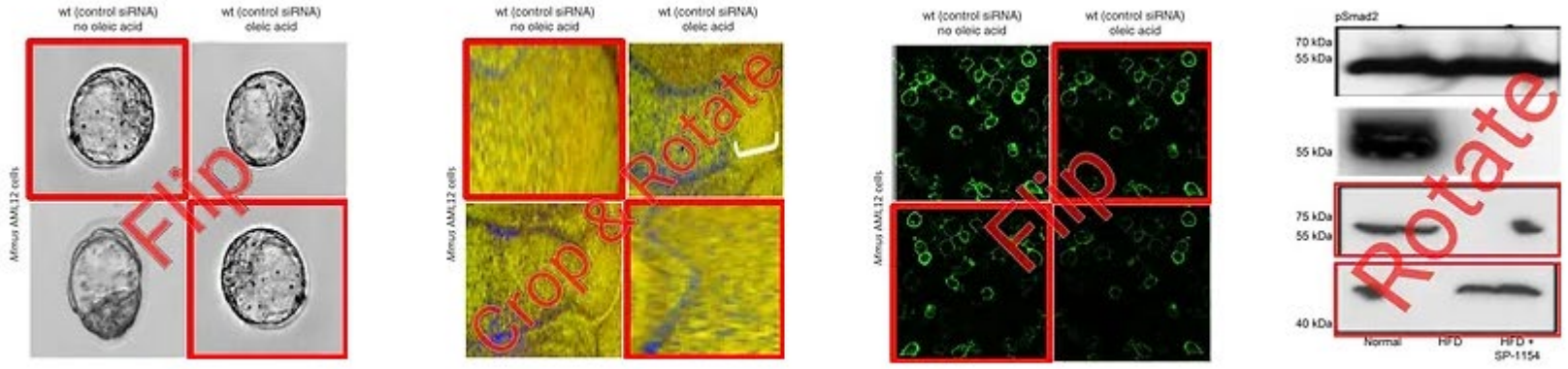
Detection of Image Manipulation

- “[American Association for Cancer Research (AACR)]publishes ten research journals and reviews over 13,000 submissions every year.
- “From January 2021 to May 2022, officials used Proofig to screen 1,367 manuscripts that had been provisionally accepted for publication and contacted authors in 208 cases after reviewing image duplicates flagged by the software. In most cases, the duplication is a sloppy error that can be fixed easily. Scientists may have accidentally got their results mixed up and the issue is often resolved by resubmitting new data.
- “On rare occasions, however, the dodgy images highlighted by the software are a sign of foul play. Four papers out of the 208 were withdrawn, and one was rejected afterwards.”
- Source: Katyanna Quach. Academic publishers turn to AI software to catch bad scientists doctoring data. The Register. September 2022. (https://www.theregister.com/2022/09/12/academic_publishers_are_using_ai; accessed on April 21, 2023)

Proofig (https://www.proofig.com/)

Ensure the highest standards of quality and image integrity in your publications

Auto detection of:
rotation, scaling, flipping, cropping, full overlap, partial overlap, cloning, and all of these issues combined



What Drives Researchers to Engage in Research Misconduct?

What Drives Scientists to Engage in Research Misconduct?

- Poor supervision
- Inadequate training
- Competitive pressures
- Personal circumstances
- Individual psychology
- Source: https://ori.hhs.gov/sites/default/files/2018-04/12_Potential_Drivers.pdf

The Toxic Lab, Replicability and Reproducibility, and Research Misconduct I

Charles Wood* wrote about two models of lab management, which "can both discourage trainees and encourage misconduct”:

> In the **executive model** of lab management, the principal investigator demands that trainees meet his or her expectations, often with a specific goal in mind. In its most toxic form, that goal can include specific experimental outcomes — so a trainee is told to do this experiment and get this particular result. [...] In the second toxic style of mentorship, the **competition model**, principal investigators give two or more trainees the same goal. The implication is that the one who completes the task first — or, more dangerously, the one who generates the data that conform best to the preconceived outcome — is the winner.

*Wood, C. Column: When lab leaders take too much control. Nature 491, 785–786 (2012). https://doi.org/10.1038/nj7426-785a (emphasis mine)

The Toxic Lab, Replicability and Reproducibility, and Research Misconduct II

The climate and environment early career researchers encounter shape their ethical behavior, according to Mumford *et al*.*

> They identified five factors of environmental experiences: "[P]rofessional leadership, poor coping, lack of rewards, limited competitive pressure, and poor career direction," and four factors of climate perception: "[E]quity, interpersonal conflict, occupational engagement, and work commitment." They found that "it appears, at least among 1st-year doctoral students, that [environmental] experience exerts stronger effects on ethical decision making than the climate of the work group."

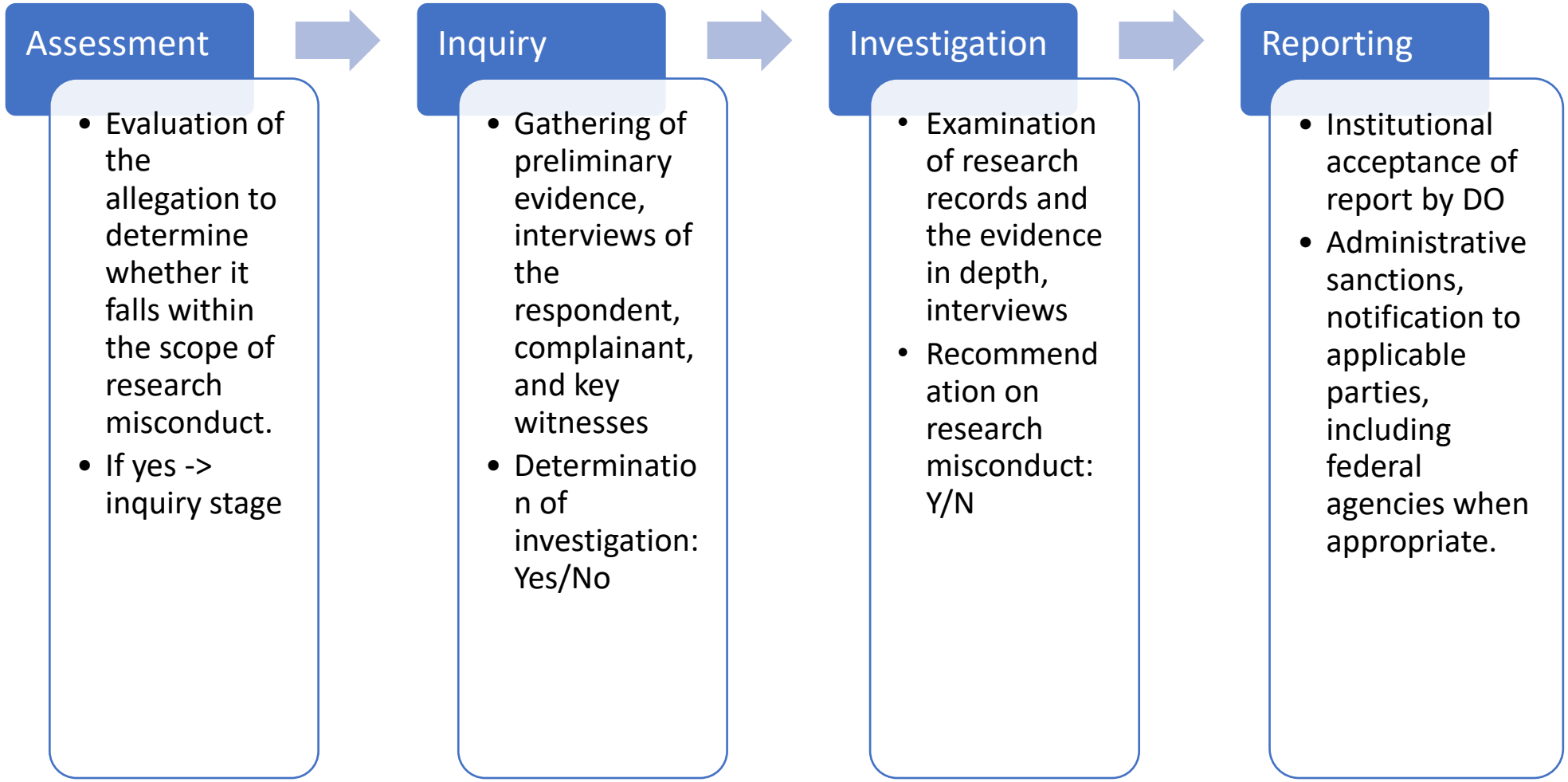
*Mumford, Michael D., et al. "Environmental influences on ethical decision making: Climate and environmental predictors of research integrity." Ethics & behavior 17.4 (2007): 337-366.

The Perils of Collaboration

- Dr. Stanley Rapoport was a section chief at the National Institutes of Health. Three of his collaborators falsified data.
  - https://retractionwatch.com/2017/07/24/nih-section-chief-19-retractions-no-longer-running-lab/
  - Dr. Rapoport no longer runs his lab at the U.S. National Institute on Aging
  - 19 of his papers have been retracted due to misconduct of co-authors
  - Source: Retraction Watch
  - Dr. Rapoport ‘s thoughts
    - “The misconduct, as I now understand it, was very technical and outside of my areas of expertise. In retrospect, I don’t think I could have spotted the misconduct earlier. Data were presented at internal meetings, when the misconduct was not identified. Basselin and Gao and Rao had PhDs and strong letters of recommendation.
    - “In these days of complex interdisciplinary research, one depends on the trustworthiness of colleagues who use methodologies with which one has no personal experience. I regret missing the falsifications by Dr. Rao…”

Research Misconduct: Process

Steps in Research Misconduct Evaluation



Source: https://uh.edu/research/compliance/research-misconduct-overview/

The Inquiry Phase

- The purpose of the inquiry is to conduct an initial review of the **available evidence** to determine whether to conduct an investigation. An inquiry does not require a full review of all the evidence related to the allegation. However, it does include the testimony of the respondent, complainant, and key witnesses to determine whether an investigation is warranted. It does not determine whether research misconduct definitely occurred or who was responsible.
- Available evidence typically includes all research data that is needed to conduct the investigation.
- **A team will come to the lab and sequester all relevant data.**

Data Sequestration and Provenance of Data

- “The term “data provenance”, sometimes called “data lineage,” refers to a documented trail that accounts for the origin of a piece of data and where it has moved from to where it is presently.
- “The purpose of data provenance is to tell researchers the origin, changes to, and details supporting the confidence or validity of research data.
- “The concept of provenance guarantees that data creators are transparent about their work and where it came from and provides a chain of information where data can be tracked as researchers use other researchers’ data and adapt it for their own purposes.”
- Source: Data Provenance. National Library of Medicine. (https://www.nnlm.gov/guides/data-glossary/data-provenance; accessed on April 22, 2023)

A Case Study

A Case Study

- “In general, older data are harder to follow compared to newer data. These truths became starkly evident to me (MEA) after a former laboratory member was discovered to have engaged in scientific misconduct.[…] However, the trail of data was incomplete and most of the publications were over ten years old. The laboratory notebooks were in hand, but his computer, left behind in a laboratory move, was lost.”
  - NIH data—stored for at least three years after competitive renewal
  - Data belong to the institution, but PI is responsible for stewardship
- Anderson, Mark E., and Stuart C. Ray. "It’s 10 pm; Do You Know Where Your Data Are? Data Provenance, Curation, and Storage." Circulation research 120.10 (2017): 1551-1554.

A Case Study (ctd.)

- “Each lab member is required to keep:
  - A bound notebook describing daily laboratory activities and referencing associated electronic files for each experiment. Records should be written with indelible ink, and errors marked, but not erased. An equivalent electronic notebook is also acceptable.
  - All raw data acquired, either physical (i.e. printouts, films, etc.) or electronic (i.e. images, patch clamp recordings, etc.).
  - Backups of all electronic data files. All data should be stored on the lab’s shared server drive or backed up on an equivalent secure server (i.e. cold storage).”
- Anderson, Mark E., and Stuart C. Ray. "It’s 10 pm; Do You Know Where Your Data Are? Data Provenance, Curation, and Storage." Circulation research 120.10 (2017): 1551-1554.

A Case Study (ctd.)

- “For each published manuscript the lead author is required to provide the following:
  - The final accepted manuscript file/s.
  - All figures (regular and supplementary) published.
  - All raw data files and calculated data files associated with each figure. The data should be organized into folders by figure.
  - A Word document (or similar) describing the files associated with each figure indicating any special software necessary for viewing files, and any additional data that are not electronic (i.e. notebooks, films, etc.). Other lab members should be able to identify the data used to create each figure.”
- Anderson, Mark E., and Stuart C. Ray. "It’s 10 pm; Do You Know Where Your Data Are? Data Provenance, Curation, and Storage." Circulation research 120.10 (2017): 1551-1554.

A Case Study (ctd.)

- “These files should be saved in a folder with the author name, journal and year published. The folder is to be stored on the lab shared server drive.
- “Upon leaving the lab, all notebooks, data backups and raw data are to remain in the lab with the lab manager.
- “Moving data to secure storage as early as possible (ideally upon acquisition) reduces risk of loss due to storage failure or omission, and minimizes additional effort when it becomes routine practice.”
- Anderson, Mark E., and Stuart C. Ray. "It’s 10 pm; Do You Know Where Your Data Are? Data Provenance, Curation, and Storage." Circulation research 120.10 (2017): 1551-1554.

Data Management Best Practices

Best Practices

- Data management plan
- File management and data formatting
- Data access
- Data analysis
- The lab meeting
- Collaborative publications
- Curating data, metadata
- Foundational data structure
  - Migrating data, controlled vocabulary,
  - Reorganizing data, cleaning data
  - Misleading results due to changing attributes during organizing data

Sound Data Management Practices Can Protect You from Research Misconduct

How to do this?

- Planning for data management
- Aligning to a foundational data structure
- Establishing collaboration and publishing norms

Data Management Plans

3 Major Components

- What are you creating / generating?
- How is it securely handled during the project?
- How is access and data integrity maintained long term?

Additional Details:
Roles and responsibilities, systems used, documentation, security

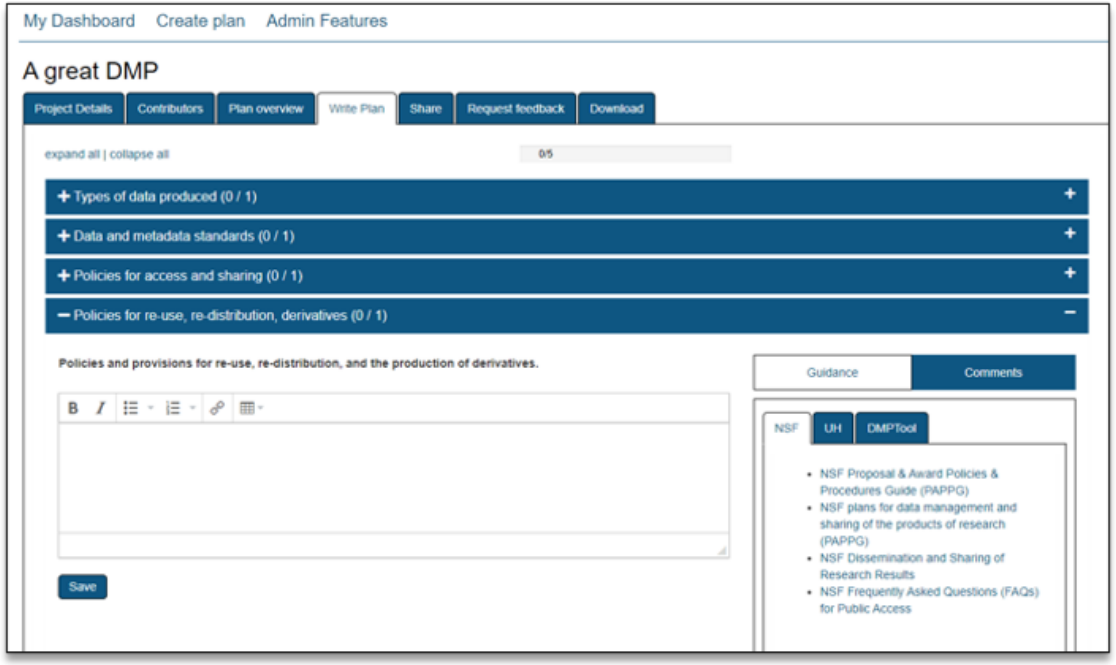
An Online Tool for Writing DMP

DMPTool

Build your Data Management Plan

https://dmptool.org

- Agency templates & guidance
- Institutional login
- Submit plans for feedback



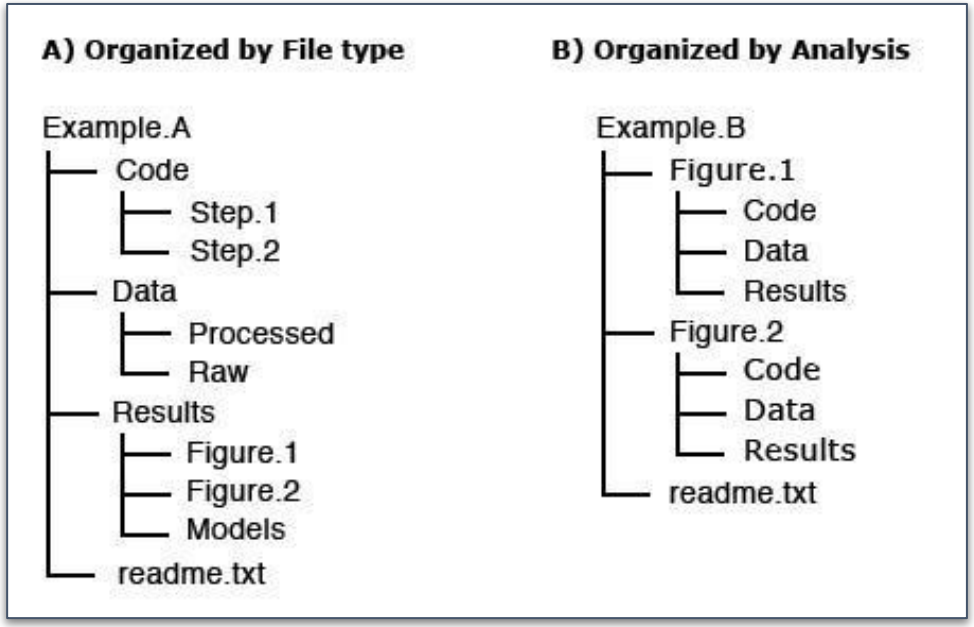
Goal: Foundational Data Structure

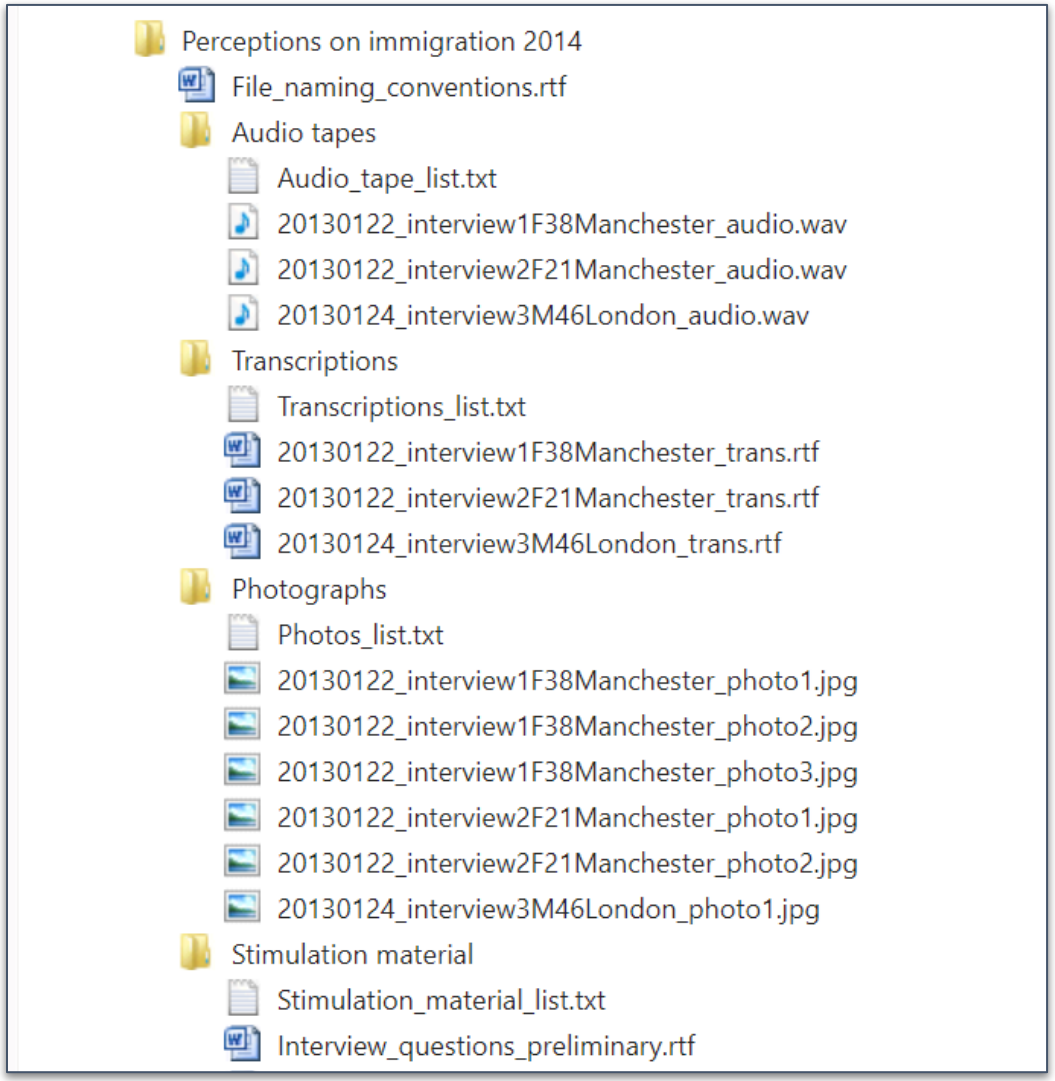
- File management and data formatting
- Data access and analysis
- Metadata
- Data curation and documentation
- Storage and version control

File-Naming Tips

- Distinctive Descriptors
- Be Consistent, Be Concise
- Use simple version control: v1, v2 etc. *(not final)*
- Use international standards for dates YYYY-MM-DD
- Avoid other characters: ~ ! @ # $ % ^ & * ( ) ` ; < > ?
- *Document your structure and convention*

File Structure and Hierarchical Organization



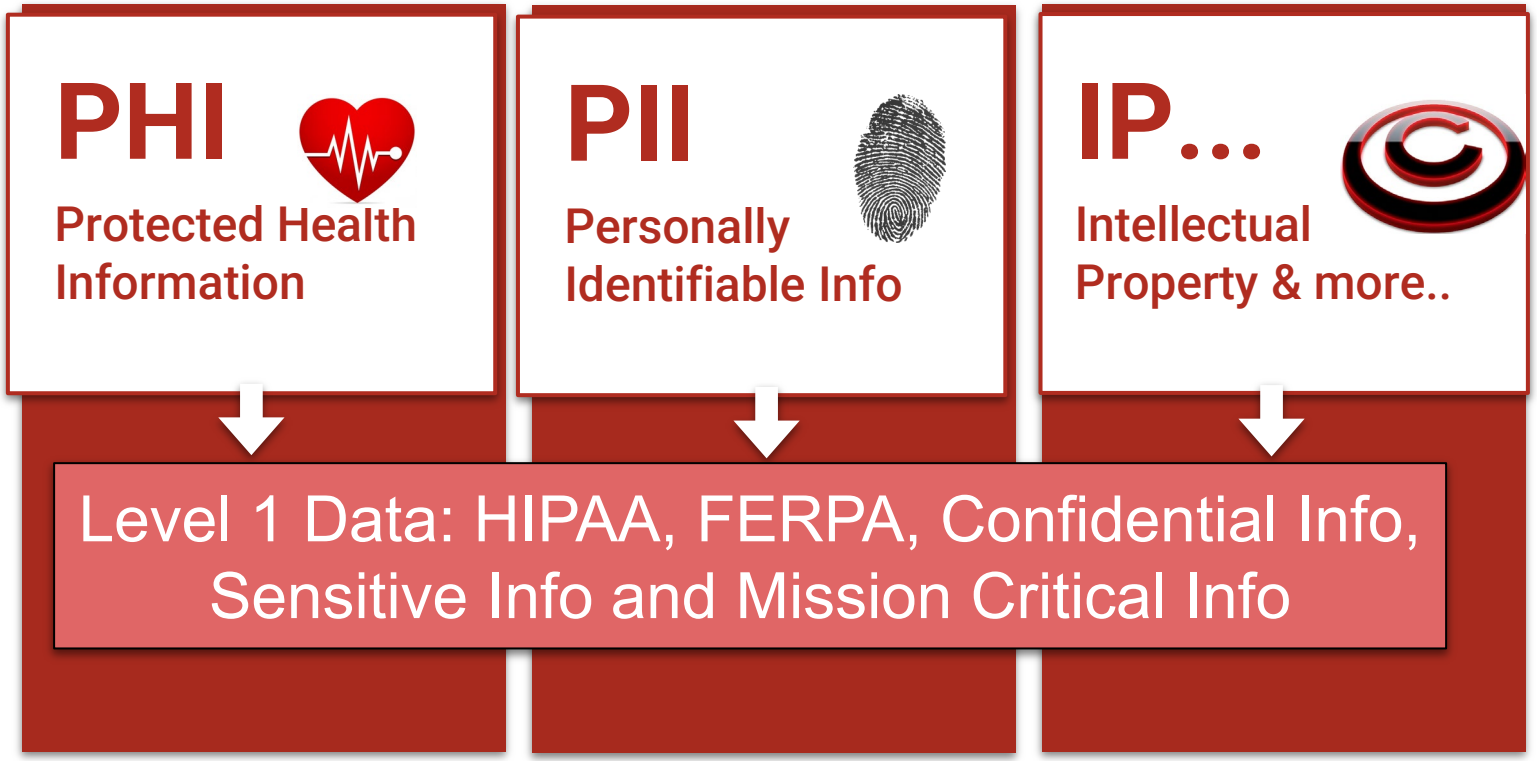


Data Formatting

- Specialized software, especially commercial, can produce file formats that are not maintained or difficult to render in the future
- Use non-proprietary formats whenever possible
- Anonymize data following institutional and disciplinary norms

Image
Source: https://blog.ironmarkusa.com/hubfs/nativefile.png

Data Access Controls



Data Analysis

- Should be a transparent process for research team and external audiences
- Document
  - Roles and responsibilities for data analysis
  - Analysis approaches and decisions
- Ensure all team members understand analysis process

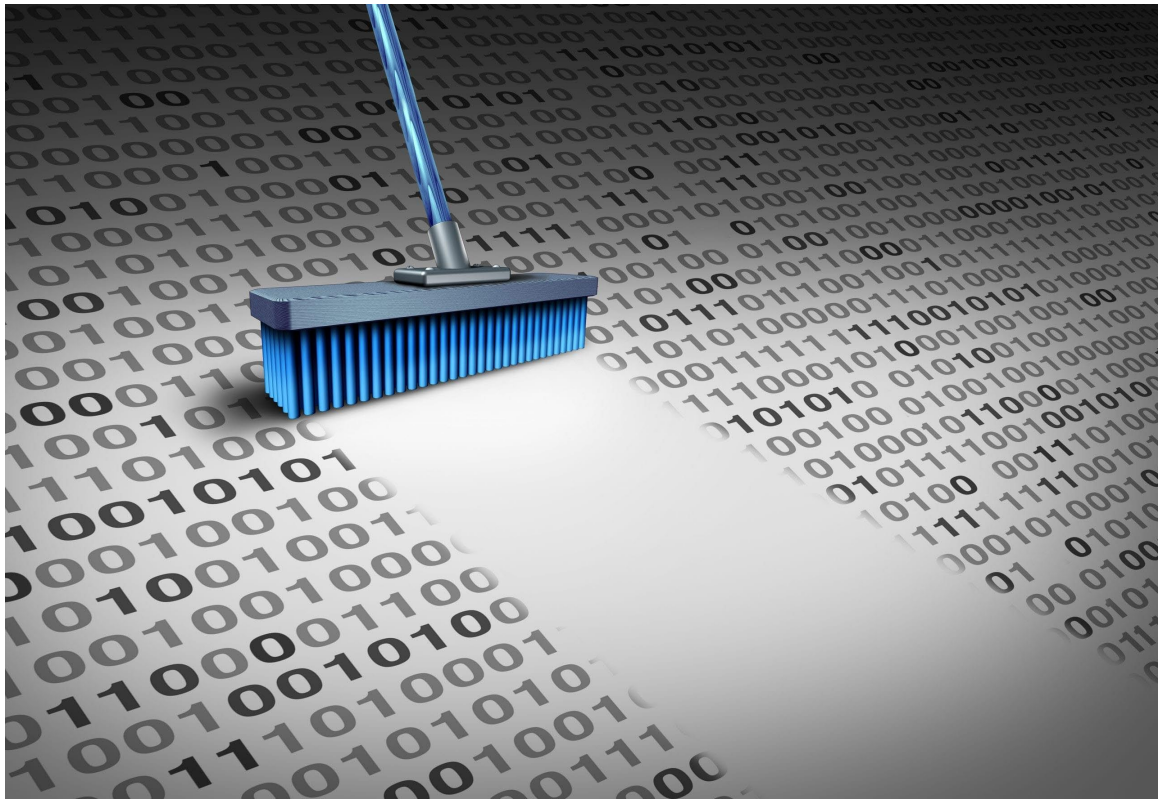
Image Source: https://img.freepik.com/free-vector/data-analysis-icons_1284-3861.jpg?w=2000

The Importance of Metadata

Descriptive information about the data, its creation, and use.
Variables, units of measure, standards, codes, procedures...

Data Curation and Documentation

- Reorganizing data
- Cleaning data
- Migrating data

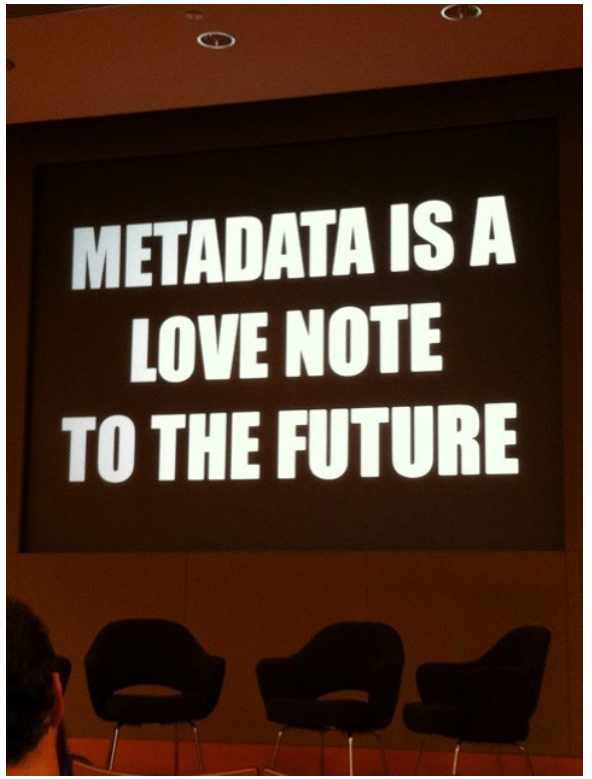


Data Curation for Reproducibility

Choosing what to make accessible...

Needed:

- File management & conventions throughout a project
- Descriptive and structural metadata (Documentation)
  - What's in the files & the relationship between them
  - Apply standards where applicable and/ or use readme files



Funder focus on sharing and access requirements is connected to quality documentation and capturing related research materials beyond the data.

Storage and Backup

3 copies - **2** different media - **1** copy off site*

Working - UH Network | Cloud - Backup 1 | External – Backup 2

*Caveats: Sensitive data / large data / costs

Raw files unable to be replicated - *treat these like gold...*

The Lab Meeting

- Create environment of constructive feedback and peer-review
- Conversations among project team can surface:
  - Unintentional misconduct
  - Errors in methods, analysis, and/or findings
- Promote healthy debates and scholarly dialogues among team

Image Source: https://journals.plos.org/ploscompbiol/article/figure/image?size=inline&id=10.1371/journal.pcbi.1008953.g001

Collaborative Publications

- Ensure project team understands the work of the group and the results it has generated
- Discuss how project team members will receive attribution for their contributions

Questions?

Contact Information

Claudia Neuhauser
cmneuhauser@uh.edu

Santi Thompson
sathompson3@uh.edu